Phishing Website Detection Using Ensemble-Based Machine Learning Approaches

Supervised By

Ms. Raiyan Janik Monir

Lecturer

Department of Software Engineering

Daffodil International University

Submitted By

Syed Naimur Rahman Rahat

ID:221-35-1019

Department of Software Engineering

Daffodil International University

This thesis report has been submitted in fulfilment of the requirements for the Degree of Bachelor of Science in Software Engineering.

DAFFODIL INTERNATIONAL UNIVERSITY

DECLARATION OF THESIS AND COPYRIGHT

Author's Full Name	:	Syed Naimur Rahman Rahat
Date of Birth	:	28/10/2001
Title	:	Phishing Website Detection Using Ensemble-Based Machine Learning Approaches
Academic Session	:	2022-2025

I declare that this thesis is classified as:

☐	CONFIDENTIAL	(Contains confidential information under the Official Secret Act 1997)*
☐	RESTRICTED	(Contains restricted information as specified by the organization where research was done)*
☐	OPEN ACCESS	I agree that my thesis to be published as online open access (Full Text)

I acknowledge that Daffodil International University reserves the following rights:

1. The Thesis is the Property of Daffodil International University.
2. The Library of Daffodil International University has the right to make copies of the thesis for the purpose of research only.
3. The Library of Daffodil International University has the right to make copies of the thesis for academic exchange.

Certified by:

Rahat

_____
(Student's Signature)

221-35-1019

_____
Student ID
Date: 27/11/2025

Raiyan

_____
(Supervisor's Signature)

Ms. Raiyan Janik Monir

_____
Name of Supervisor
Date: 27/11/2025

APPROVAL

APPROVAL

This thesis titled on **"Phishing Website Detection Using Ensemble-Based Machine Learning Approaches"**, submitted by **Syed Naimur Rahman Rahat (ID: 221-35-1019)** to the Department of Software Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of Bachelor of Science in Software Engineering and approval as to its style and contents.

BOARD OF EXAMINERS

Fazla Ealhe

Chairman

Dr. Fazla Ealhe
Assistant Professor & Associate Head
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Marzia

Internal Examiner 1

Dr. Marzia Ahmed
Assistant Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Internal Examiner 2

Dr. Shabnom Mustary
Assistant Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Internal Examiner 3

Md. Rajib Mia
Lecturer (Senior Scale)
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

External Examiner

Mohammod Abul Kashem, PhD
Professor
Department of Computer Science and Engineering
DUET, Bangladesh

Phishing Website Detection Using Ensemble-Based Machine Learning Approaches

Syed Naimur Rahman Rahat

ID:221-35-1019

Bachelor of Science

DAFFODIL INTERNATIONAL UNIVERSITY

SUPERVISOR'S DECLARATION

I hereby declare that I have reviewed this thesis entitled "**Phishing Website Detection Using Ensemble-Based Machine Learning Approaches"**, and in my opinion, it is adequate in terms of scope and quality for the award of the degree of Bachelor of Science in Software Engineering.

(Supervisor's Signature)

Full Name : Ms. Raiyan Janik Monir

Position : Lecturer, Department of SWE, DIU

Date : 22 December 2025

STUDENT'S DECLARATION

I confirm that the piece in this thesis is based on my own writing with the exception of quotation and reference that have been discussed. I also confirm that it was not previously and concurrently registered at Daffodil International University or other institutions at any other degree.

Rahat

(Student's Signature)

Full Name : Syed Naimur Rahman Rahat

ID Number : 221-35-1019

Date : 22 December 2025

Phishing Website Detection Using Ensemble-Based Machine Learning Approaches

Syed Naimur Rahman Rahat

ID:221-35-1019

Thesis submitted in fulfilment of the requirements
for the award of the degree of
Bachelor of Science

Department of Software Engineering

DAFFODIL INTERNATIONAL UNIVERSITY

DECEMBER 2025

ACKNOWLEDGEMENTS

I would like to take this opportunity to list my heartfelt thanks for the help and encouragement from all the people who helped me finish this thesis titled “Using Machine Learning Model with Ensemble Technique to Detect Phishing Websites”. I am indebted primarily to Ms. Raiyan Janik Monir, Lecturer, Department of Software Engineering, Daffodil International University (DIU) for her supervision, guidance and encouragement during the work on this study. Her valuable suggestions, guidance and expertise have been absolutely vital to the success of this work. I am also thankful to the Department of Software Engineering, Daffodil International University for providing a conducive academic atmosphere, facilities and resources which enabled me to carry out this research. I owe a heartfelt debt to all my friends and colleagues for their valuable comments, cooperation and encouragement throughout the research work. Last but not least, my sincere thanks go to my dear family members for their whole-hearted love, moral support and encouragement. It’s that unwavering faith in me that has given me the strength to carry on and to push myself forward when the going gets tough.

DEDICATION

This work is dedicated to my beloved parents, whose boundless love, support and sacrifices have formed the bedrock of my accomplishment. Their undying support, prayers and believing in my capabilities pushed me to be at my best despite the obstacles. Dedication I further dedicate this thesis to my dear supervisor, Ms. Raiyan Janik Monir, for her sincere guidance, support and priceless supervision during the period of this research. Without their love, support and inspiration this would not have been possible.

ABSTRACT

Phishing is relatively one of the most common and dangerous cybersecurity threats in existence and convinces users to expose sensitive information by using fake websites. These attacks have dynamic structures and advanced obfuscation techniques that render traditional detection methods ineffective against them. A Pragmatic Analysis on Detection of Phishing Based on URL · This work advocates a URL phishing detection framework based on machine learning that can differentiate the legitimacy of the URLs in high accuracy. To overcome the imbalance of the data, balanced data was created by using the SMOTE technique and then three efficient models were implemented and trained, namely, LightGBM, Random Forest and XGBoost. The strengths of these base classifiers were used to combine them and create a hybrid stacking ensemble model called HyPhish-Net to achieve better detection performance. To enhance interpretability and enhance performance, feature selection as well as correlation analysis were conducted comprehensively. Based on the experimental results shown, all base models give a flexibly robust accuracy but the highest accuracy features score in the proposed HyPhish-Net model is 98.6%, of precision of 0.987 and recall of 0.985. The assessment proved that the model could drastically decrease false footage and false footage, assuring the high reliability of phishing detection. The performance metrics, including confusion matrices and ROC analysis, showed excellent generalizability of HyPhish-Net across training and testing data.
A robust equilibrium between sensitivity and specificity, outperforming the best individual models on all the key indices of its performance, was attained by the system. The proposed model can be deployed in browsers, email systems and cybersecurity infrastructures because of its robustness and scalability. Ultimately, the study concludes that if intelligently implemented, ensemble-based learning provides an effective solution for automated, intelligent phishing detection. In summary, this thesis helps us with implementing more secure and adaptive detection mechanism to protect users from real world digital environments.

Keywords: Phishing Detection, Machine Learning, LightGBM, Random Forest, XGBoost, Ensemble Learning, Stacking Model, HyPhish-Net, Cybersecurity, URL Classification, SMOTE, Feature Correlation, Precision, Recall, Accuracy.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	Full Form
AI	Artificial Intelligence
AUC	Area Under the Curve
CNN	Convolutional Neural Network
DNS	Domain Name System
DT	Decision Tree
GBM	Gradient Boosting Machine
GA	Genetic Algorithm
HTTP	Hypertext Transfer Protocol
HTTPS	Hypertext Transfer Protocol Secure
KNN	K-Nearest Neighbors
LR	Logistic Regression
ML	Machine Learning
NB	Naïve Bayes
ROC	Receiver Operating Characteristic
RF	Random Forest
SVM	Support Vector Machine
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
URL	Uniform Resource Locator
XGBoost	Extreme Gradient Boosting
HyPhish-Net	Hybrid Phishing Detection Network (Proposed Model)

CHAPTER 1

INTRODUCTION

1.1 Introduction

Phishing is one of the most common and damaging forms of cybercrime that targets individuals, organizations, and even entire critical infrastructure for financial gain or unauthorized access. Although the goal of generic phishing attacks is to compromise a lot of victims, tailored phishing attacks are sophisticated and target specific people, sets of people, or organizations. Due to their unique or time-based nature, these targeted phishing attacks usually bypass common detection methods like phishing blacklists [1]. Phishing attacks are present in virtually every digital space and service means they are everywhere: You have phishing in email, social media, instant messaging is all a baited hook that other forms exist for the scammer to steal. Unlike the phishing reported email service impersonation and communication trust attacks exploits the fact that many people use these services for both personal and business communication [2,3]. These are social engineering hacks, basically trying to convince someone to give away their security level permission. Once considered basic email scams, now these attacks have become sophisticated multi-step actions, able to manipulate even seasoned users and avoid high-end security systems. As a consequence of a phishing attack, there can be major loss of money, damage to reputation, identity theft, and even legal action. It is more commonly linked to data breaches, ransomware infections, and other criminal cyber acts. These threats are likely to spill over into other areas of security and impact individuals and organizations. Cybersecurity experts and security researchers are constantly creating and making efforts to implement solutions to identify, combat, and mitigate phishing [4]. Phishing detection based on machine learning can detect zero-day phishing attacks even before they can be blacklisted. But all recent advances empower only a patchwork of initiatives due to a critical gap [5]. Though current phishing detection models perform Capably on English or other high-preferred languages, the performance is sub-optimal when it come to able webpages in minor languages [6]. The major goal of this thesis is to create and test strong machine learning–based models for the precise identification of phishing websites and emails. The study experiments with several algorithms including LightGBM, Random Forest and XGBoost and with a proposed stacking ensemble model HyPhish-Net to find the best solution to classify phishing and benign instances.

1.2 Background Study

As the world evolves through the digital technologies of the 21st century, they are becoming more integrated, making cybersecurity an unavoidable global concern. As reliance on online services, communication platforms and digital transactions increases, the frequency and sophistication of cyber threats has also grown [7]. Of these, phishing has become one of the most frequent and most vicious types of cybercrime. Phishing is derived from fishing; attackers attempt to trick victims into throwing out their private information. When cybercriminals impersonate trusted entities like banks, government organizations, or social media platforms to trick users into disclosing sensitive data. Identity theft, money loss, and loss of privacy — all as victims unintentionally reveal personal information, passwords, and bank details. Such attacks on individuals are also seen in the disorganization of organizations and financial institutions. With the innovation of phishing and cyber professionals coming up with new ways to attack and breach the system, it has become crucial to strengthen the awareness about cybersecurity and developing response strategies and mechanisms to protect your target users in the digital era.

1.3 Motivation

Phishing is one of the oldest and the most evolving cyber threats in the landscape of modern digitalization. With cybercriminals continually improving their strategy, traditional detection systems used in the past have proven unsuccessful in keeping up with the change in attack behavior. Phishing incidents have become more prevalent and sophisticated which necessitates dynamic and smart defense approaches. There is now an urgent need for an automated detection method that can also handle big data, as well as differentiate between normal and abnormal users accurately. Improvements like these are critical not only for the safety of our users, but the durability of the information ecosystem itself.

1.4 Problem Statement

Phishing attacks evolve faster than we think, targeting humans and systems to create situations where a loss is inevitable to ensure minimal damage, Phishing attacks exploit humans and systems to ensure loss. Detection systems in use today are mostly based on manual extraction of features and rules and this makes them a bit grey and quite ineffective against new and unseen attacks. These restrictions are responsible for high rates of false-positives and limited flexibility in practical applications.

The rapid growth of online data has created the demand for a scalable [1], automated [2] and precise phishing detection framework which learns evolving trends from several datasets. Tackling this problem is essential for enhanced detection capability, shortened response duration, and a more secure cyberspace for users globally.

1.5 Research Objective

This research study is mainly focused on the framework of phishing detection via machine learning with different algorithms. The work compares several models such as LightGBM, Random Forest and XGBoost for classification performance in phishing task. It intends to build a stacking ensemble model HyPhish-Net configures which combines the properties of base learners with a meta-classifier allowing for a higher predictive accuracy. Another task is to use SMOTE to resolve class imbalance and allow for balanced learning leading to improved generalization. It also aims to optimize the model using hyperparameter tuning to improve accuracy and minimize false positives. Finally evaluate model performance using several metrics: Accuracy, precision, recall, F1, ROC-AUC and the Anti-Phishing Score, which we propose. A harvesting, automated, and scalable phishing detection model is the final output to provide the answer to find malicious entities with high accuracy in real-time with low error rate.

1.6 Purpose of this Research

This research aims to provide support for robust cybersecurity by creating a smarter Automatic system for phishing detection with a higher detection rate. The proposed framework in this study offers an adaptive solution to improve phishing detection and response efficiency by minimizing detection errors, as such threats keep evolving and taking advantage of human vulnerabilities. The research applying the machine learning and ensemble learning idea to improve the accuracy and reliability of phishing detection. Additionally, the objective also includes raising awareness over active cyber hygiene that can protect users, businesses, and banks from cyberattacks. By evaluating and comparing with a systematic lens, this work aims to pave a critical role for the advancement of phishing detection and cyber defense solutions in the real world.

1.7 Organization of the Thesis

Chapter 1 Introduction: Introduces research background, statement of the problem, objectives, motivation, purpose and significance of study.

Chapter 2 Literature Review: Reviewed previous work on the dengue detection and machine learning techniques covering their limitations and why this type of model possesses a practical value.

Chapter 3 Methodology: Presents the dataset, preprocessing, model construction, and model specifications of proposed HyPhish-Net hybrid ensemble.

Chapter 4 Results and Analysis: Presents the experimental results, model performance evaluation, comparison with other techniques and visual interpretation of results.

Chapter 5 Conclusion of the study: Presents the general conclusion, summarizes with findings obtained, identifies issues relating to the conclusion, concludes the study and recommends possible areas for future investigations.

CHAPTER 2

LITERATURE REVIEW

2.1 Overview

Phishing is one of inducing the user to provide sensitive information in fake websites that is common threats on cyber. Conventional blacklist-based and heuristic-based approaches are reactive in nature and instantly may not recognize newly appeared phishing sites. Accordingly, more and more efforts have been devoted to using machine learning (ML) and ensemble learning techniques for identifying phishing websites by combining URL, domain, as well as content characteristics. Ensemble methodologies like Random Forest, AdaBoost, Gradient Boosting and XGBoost have attracted attention for their capability to aggregate many weak learners with high accuracy and robustness. This subsection describes the related works of ML and ensemble for phishing website detection focusing on used techniques, datasets and results.

2.2 Previous Work

Zhang et al. (2007) presented CANTINA, a content-based phishing detection technique that applies TF-IDF for keyword extraction and heuristic checks such as domain age and URL length. The model reached about 97% exactitude of label on phishing sites, but this measure was compared with a 6% false positive rate. Their work demonstrated the necessity of both content and heuristics features for accurate detection. Fette et al. (2007) proposed PILFER: an email-based Phishing detector which is based on 10 hand-crafted features like IP-based URLs and form actions. With machine learning classifiers, the system was able to classify phishing mails with over 96% precision and proved more efficient than rule-based spam filters. This work set precedence for ML based phishing detection systems. Abu-Nimeh et al. (2007) performed a comparative analysis of several classification algorithms such as Naïve Bayes, Decision Tree, Random Forest, SVM and Logistic Regression over phishing emails data sets. No single model was found to consistently outperform the other models, implying that ensemble approaches can be adopted for making these predictors more robust and accurate in detecting phishing. Dietrich et al. (2009) presented an amalgamated scale bale machine learning methodology that leverages lexical and host-based features together in order to characterize malicious URLs. The authors obtained a 95–99% accuracy and proved the superiority of ML against conventional blacklists. This early work laid the groundwork for automated phishing detection based on URLs.

Mohammad et al. (2012) built the popular UCI Phishing Websites Dataset comprising 30 URL, HTML, and SSL based features. They have implemented the Decision Tree and Random Forest were able to get an accuracy of 95 per cent. It set a milestone for future studies in phishing detection. Nagunwa et al. (2019) proposed a hybrid feature engineering method that realized integration of URL features, content-based features and domain-based features for zero-hour detection of phishing website. Their approach attained 96% accuracy, demonstrating the importance of combining several feature sets for early phishing detection. Subasi and Kremic (2020) compared AdaBoost with Multibooting for phishing site identification with UCI dataset. AdaBoost obtained an accuracy of 95.2% and Multibooting reached 96.7%, which suggests that boosting ensemble methods improve more than individual models.

Al-Hadhrami et al. (2021) designed a fine-tuned stacking ensemble model with Genetic Algorithms to optimize base learners (Random Forest, AdaBoost, and XGBoost). The overall accuracy for the ensemble reached 98.58%, which is superior to single models. The work demonstrated the importance of meta-learning when applied to phishing websites detection.
Li et al. (2021) proposed another level stacking ensemble model using Genetic Algorithms for hyperparameter tuning of the base classifiers. Their method achieved an accuracy of 97.4% with high precision and recall, demonstrating the usefulness of applying optimization to ensemble classifiers. Sun et al. (2023) applied six traditional ML methods: Logistic Regression, KNN, Naïve Bayes Random Forest, SVM and XGBoost for Phishing detection. XGBoost was found to be the best out of these with accuracy of 96.2% and highest distinction ratio of 99.4% that confirms it is indeed a better gradient approach for detection of phishing websites.

Basnet et al. (2023) proposed a hybrid phishing check system, utilizing URL and HTML-based features combined with Random Forest and XGBoost. They reported an accuracy of over 98% showing promising results that combination of multiple feature pooling algorithms with ensemble classifiers were very successful. Alhuzali et al. (2024) proposed deep learning models for detecting phishing email. CNN and LSTM have been experimented over a relatively large email corpus and the performance of 96.7% is achieved, proved that DNN based architecture can outperform the traditional ML-based systems written for text-based phishing detection.

Raza et al. (2024) that provides a review of the evolution from classical ML to ensemble learning, deep learning methods in phishing detection. They also highlighted the performance of ensemble models such as Random Forest and XGBoost, whose likelihood of being manipulated shifts alongside phishing tactics. PhishNet The work in [16] established an XGBoost-based detector that used URL structure, domain age and WHOIS information as feature vectors to determine the class of websites. It demonstrated 98.9% accuracy, low false positive rates and indicates that practical in-situ ensemble tools might be operationally feasible. [IJNRD 2024] performed train/test on Gradient Boosting classifiers with over 11055 URLs dataset and obtained accuracy of 97.4% along with precision, recall value as >0.97. This indicated that gradient boosting could generalize well on phishing datasets.

IJRAR (2024) presented a Gradient Boosting Classifier for website authentication with 98.2% accuracy and highlighting its feasibility for Realtime phishing detection system. Mishra et al. (2025) designed an XGBoost model using Bat Algorithm for hyperparameter tuning. Their vocabulary obtained accuracy of 99.3% showing the way how metaheuristic tuning improves theories capacity for detection Ing phishing websites. Kaur et al. (2025) provided an up-to date literature review on and comparison between ML and ensemble detection algorithms for phishing. They have found that ensemble methods provide better results compared to single models when coupled with state-of-the-art optimization and feature selection techniques.

CHAPTER 3

METHODOLOGY

3.1 Overview

In this Thesis, we propose a phishing detection framework, which represents a more structured and data driven methodology. Our first step involves gathering a large phishing dataset covered by genuine as well as phishing samples. In the data preprocessing phase, we remove the unnecessary attributes, the missing values are solved, and we convert categorical features into numerical features to be more suitable for the machine learning models. To overcome the class imbalance problem, applies the SMOTE (Syntactic Minority Over-sampling Technique) method, where phishing and non-phishing data are equitably represented during the training phase. Once the dataset is preprocessed, we split the data into training and testing subsets using the train-test split method to judge how well the model generalizes. Three of the best performing models—LightGBM, Random Forest, and XGBoost—are trained individually to learn patterns and classify data properly. They are then integrated using a stacking ensemble model called HyPhish-Net, where the outputs of the base models are fused by a Logistic Regression meta-classifier to enhance the detection performance. Four performance metrics are used to evaluate models, including Accuracy, Precision, Recall, F1-Score, ROC–AUC, and a newly introduced Anti-Phishing Score score in order to provide a thorough assessment. Hence, this methodological approach is scaling, flexible, and resilient in terms of phishing detection in real-world cybersecurity context.

3.2 Workflow

So, what you see above is a 4-step machine learning pipeline (probably similar to your thesis) slightly customized for building a machine learning classification model using ensemble methods like Random Forest, LightGBM and XGBoost. Here is a compact summary in 7 lines as you asked:

Step 01 – Data Preparation: Take care of missing data, clean columns and change the type of variables.

Step 02 – Balancing the data: Use SMOTE to deal with imbalanced class.

Step 02 - Train-test Split the data set to be a ratio of 80 for training and 20 for testing.

Step 03 - Model Fitting: Fit the Random Forest, LightGBM and XGBoost models.

Step 03 - Deploy HyPelsi-Net (it is a private hyperparameter tuning network).

Step 04 –Evaluate Model: Check the Accuracy, Precision, Recall and F1-Score.

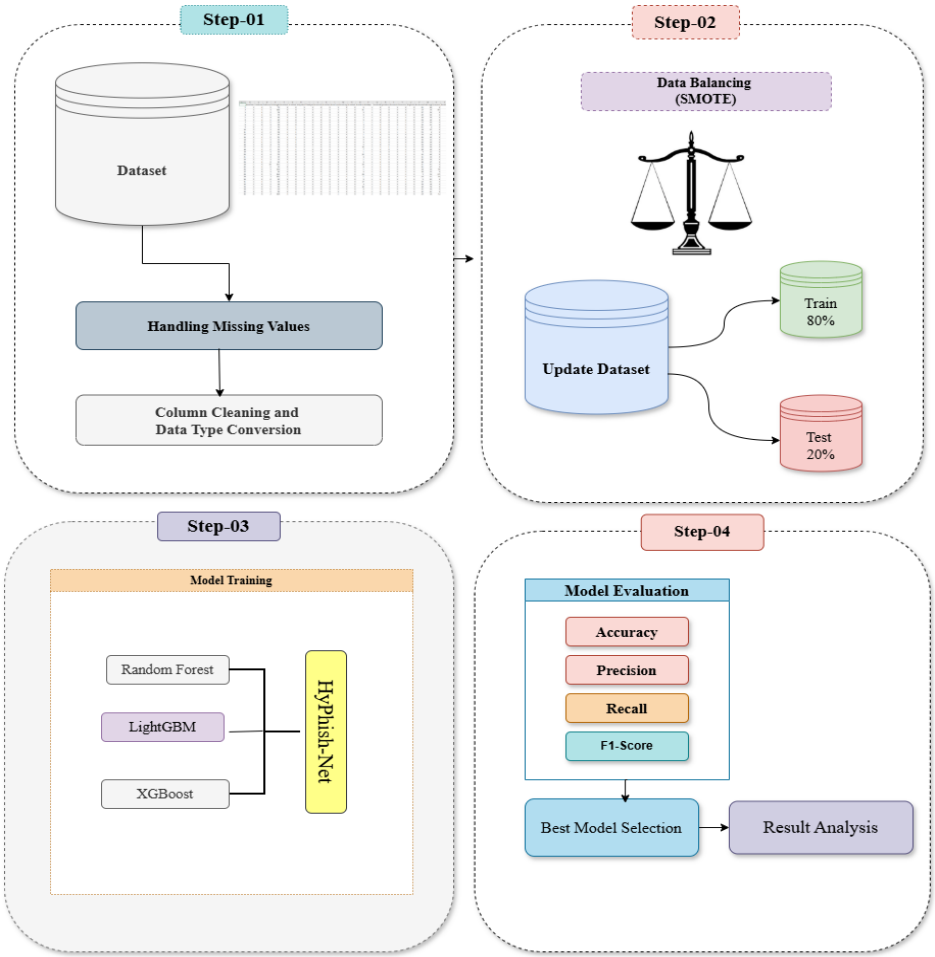

Figure 3.1: Workflow of the machine learning model development pipeline.

3.3 Dataset Description

Dataset In this work, we use a dataset comprising of 10,000 URL instances (phishing/legitimate) each of which is represented by 50 features representing structural, lexical and behavioral features. These attribute types take into account structural features (e.g., NumDots, SubdomainLevel, PathLevel and UrlLength), so- ciograms identifiers that capture varying symbol patterns like NumDash, AtSymbol, TildeSymbol and NumUnderscore reflecting obfuscation tactics prevalent among phishing links.

We use indicative attributes like NumDashInHostname for host name manipulation cues, and features like IframeOrFrame, PopupWindow, RightClickDisabled, Missing Title for Webpage behavior. Other signs indicate deceptive practices, like FakeLinkInStatusBar and SubmitInfoToEmail, while resource-based characteristics such as PctExtResourceUrls, AbnormalExtFormAction, and PctExtNullSelfRedirectHyperlinks inform about external call pattern. The dataset also contains runtime-transformed versions of key metrics, such as UrlLengthRT and SubdomainLevelRT, which boost model learning. The target variable, CLASS_LABEL, is binary (1 for phishing and 0 representing legitimate), we can use the dataset to construct/evaluate ML models for detecting phishing sites.

3.3.1 Dataset Source

The dataset used in this study, Phishing_Legitimate_full.csv, originates from publicly accessible repositories that compile real-world phishing and legitimate for research. The features consist of automated -parsing tool and webpage analysis scripts made by previous researchers that have been incorporated into the data. The original data set was created for academic research concerning phishing attacks and is a common benchmark for classification model experiments. It offers the clean, preprocessed attributes that make us less dependent on manual feature engineering. The file was accessed from open research repositories, which enable free access and use of the resources. The version I've used in this research is exactly the one available from the data set repository, without any change. In the end, both reliability and reproducibility as well as relevance for machine-learning-based cybersecurity research are guaranteed by the dataset source.
Link - https://www.kaggle.com/datasets/evilspirit05/phising-data/code

3.3.2 Applying SMOTE

Class imbalance can greatly affect the performance of the model in any classification task, even more in the cybersecurity and phishing detection domains. Despite having 2700 legitimate (class 0) URLs for 1 phishing (class 1) the original dataset, thus resulting in a high-class imbalance. Since the data distribution is not balanced, our model may tend to bias for the majority class and generate an accuracy score with a high number but with a poor recall score for the phishing URLs which is the most essential to be predicted. In order to synthesize this,

the Synthetic Minority Over-sampling Technique (SMOTE) was used. SMOTE, or Synthetic Minority Oversampling Technique, is a sophisticated data-level balancing tool which generates overlapping synthetic sample points for the minority class instead of replicating records. It generates new, realistic examples by calculating nearest neighbors of minority samples in feature space and creating data points in between along line segments connecting the neighbors. This not only balances the dataset in numbers but also maintains the data distribution to help the model learn generalized boundaries between classes.

Table 3.1: Balanced dataset after applying SMOTE.

CLASS LABEL	MEANING	NUMBER OF RECORDS
P (1)	Phishing	5000
N (0)	Legitimate	5000
TOTAL		10000

The SMOTE explained and what happened here as a result, the final dataset will be perfectly balanced, which will have 5,000 phishing and 5,000 legitimate samples total of 10,000 records. This way, the machine learning model will be trained with equal prominence of both classes. Consequently, various metrics such as recall precision and overall generalization power of the model improve significantly mainly for the phishing class, which is usually very poor. This balance is visually confirmed in the bar chart from Table 3.1, numbers of samples for each class legitimate and phishing are shown. If you follow the visualization, you will see the data imbalance problem has been practically alleviated. The summarized numerical distribution of classes is displayed down below Table 3.1, indicating that each class is represented equally.

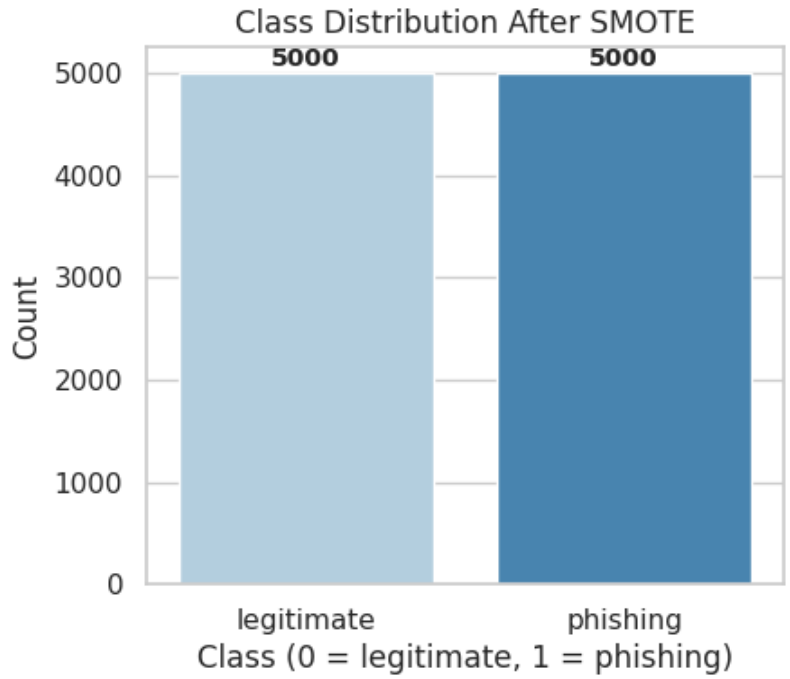

Figure 3.2: Balanced dataset after Applying SMOTE

3.3.3 Correlation Matrix

A correlation matrix is a very reliable plot that shows the relationship (positive or negative) between various numerical features in the dataset. Every cell in the matrix contains a correlation coefficient between –1 and +1, where values close to +1 indicates a strong positive correlation (the features increase together), –1 strong negative correlation (one increases while the other decreases), and around 0 indicates they are independent and have no linear relationship. A heatmap is depicted in color (more dark = stronger correlations, more bright = lower ones) in Figure 3.2, and shows these relations. With this matrix - we will be able to find redundant, or highly correlated features that could lead to additional issues like multicollinearity that can hamper prediction performance and interpretation.

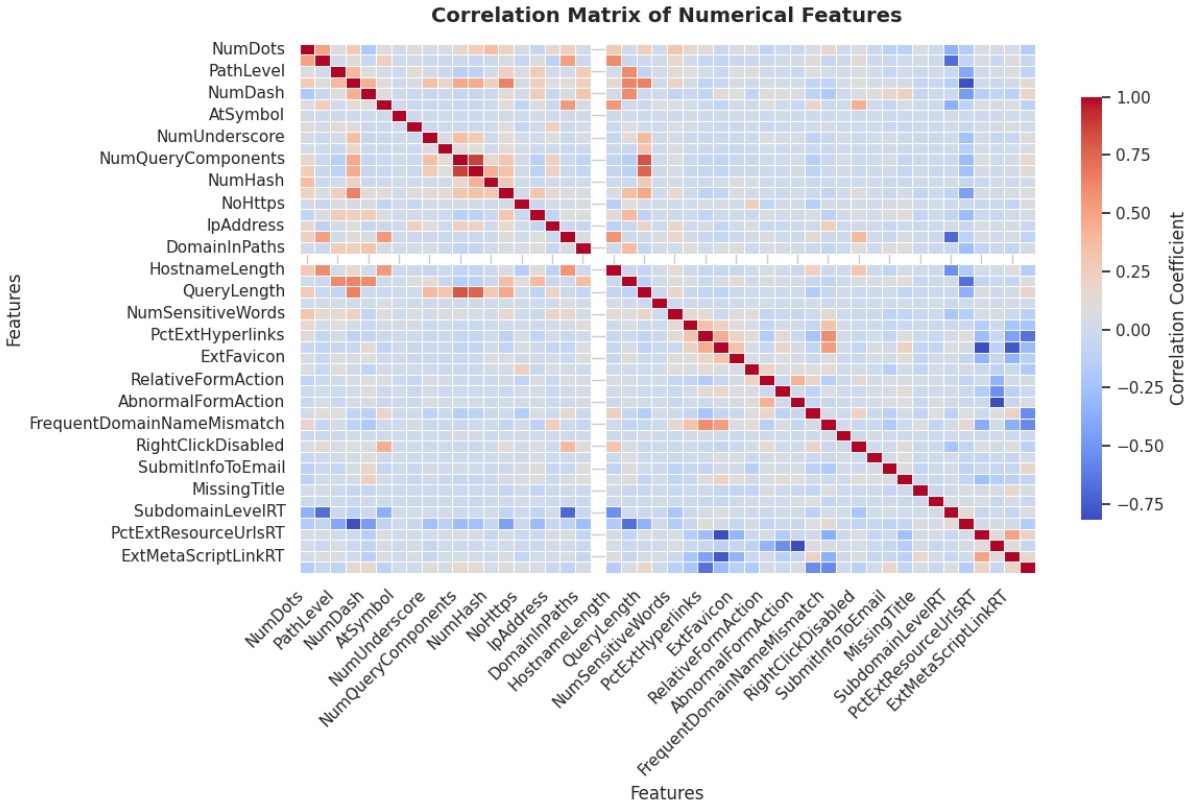

Figure 3.3: Heatmap of feature correlations

3.3 Data Split

The figure 3.3, below, where we can see the number of legitimate and phishing samples before the dataset split and after it (to create training and test data). Originally there were an equal number of 0 (legitimate) and 1 (phishing) or 50%–50% in both classes (5,000 odd 00 and 5,000 odd 01). This balance was maintained during the train-test split; there are 4,000 samples per class in the training set, and 1,000 samples per class in the testing set. By uniformly performing the train test split, both the subsets will have the same proportions of classes thus ensuring that no class is favored more than the other class while training and evaluating the model. Equal representation across splits helps to support good model performance and fair evaluation metrics.

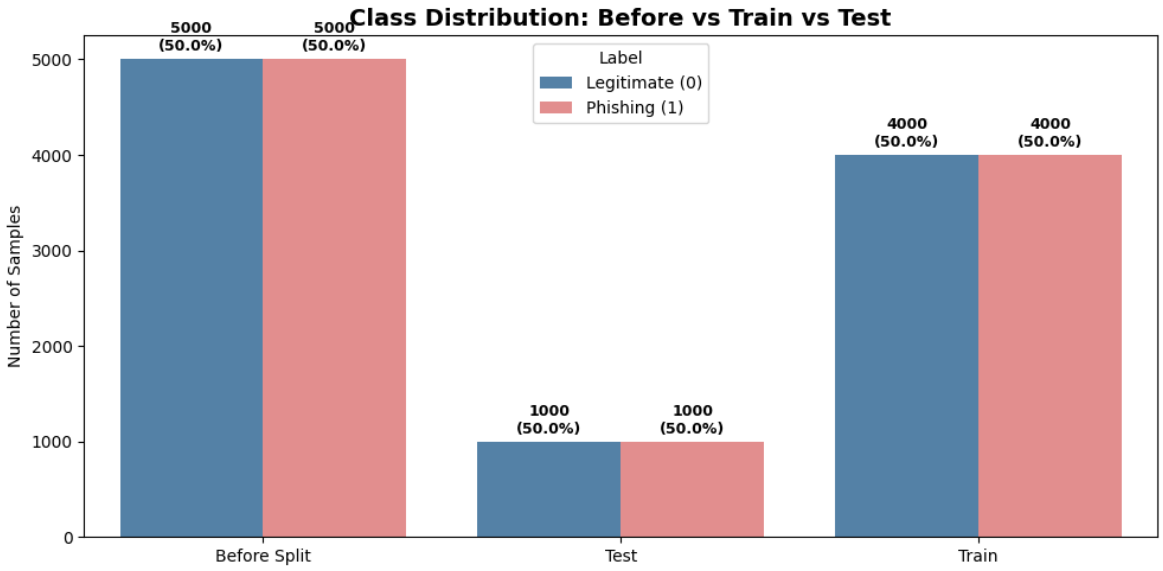

Figure 3.4: Visualization of Data Split

3.4 Training & Evaluation

Accuracy: Accuracy is more of a measure of models fit (calculating the percentage of samples that are predicted correctly in all predictions).

$$\textbf{Accuracy} = \frac{(TP+TN+FP+FN)}{TP+TN} \qquad 3.1$$

Precision: Measures the ratio of truly predicted cases to the total predicted positives.

$$\textbf{Precision} = \frac{TP}{TP+FP} \qquad 3.2$$

Recall: Recall is the portion of all positive cases that models able to find out.

$$\textbf{Recall} = \frac{TP}{TP+FN} \tag{3.3}$$

F1 Score: The F1-score is the harmonic mean of the Precision and Recall. It yields a trade-off between these two indices, particularly when data is imbalanced.

$$\textbf{F1} = 2 * \frac{Precision * Recall}{Precision+Recall} \tag{3.4}$$

3.5 Model Architecture

The model architecture includes three versions constructed using some advanced machine learning algorithms, namely, Light Gradient Boosting, Random Forest, and Extreme Gradient Boosting. The models are trained on the balanced dataset where the frequency of phishing and legitimate URLs is the same. The models are of gradient boosting nature since they learn complex patterns of features of URL stimuli and their patterns predicting one of the two classes of URLs. Gradient boosting principals are also exercised, yet in different ways Random Forest learns multiple decision trees, each of which is built on a random sample of features; LightGBM and XGBoost are similar but there are differences in the way the training is performed and the learning rate. Such an architecture provides robust learning and outstanding accuracy, which leads to reliable performance based on all evaluation metrics.

3.5.1 Light Gradient Boosting

The Light Gradient Boosting Machine (LightGBM) is a fast, tree-based algorithm that builds models sequentially with gradient boosting. It uses leaf-wise splits instead of level-wise, which results in a more accurate and efficient algorithm for large datasets. LightGBM keeps track of features, working for both categorical and numerical features, and is highly efficient for speed

and memory. Under this framework, LightGBM learns complex interactions between input features like URL length, presence of certain special characters, and domain-specific features and predicts whether a URL is a phishing or legitimate URL. It also minimizes binary log loss and can be learned in parallel, making this algorithm scalable for larger scale phishing detection tasks. Its performance using accuracy, precision, recall and ROC-AUC is also evaluated with it achieving good generalization in unseen data.

3.5.2 Random Forest

The Random Forest classifier consists of a number of decision trees trained on random data and feature subsets. Every tree casts a vote for a class, and the mode of their votes is taken as the output. In this way this randomness helps to avoid overfitting and leads to better robustness of the model. Random Forest performs great on tabular, structured data while the relationship between the URL-based features may be non-linear. The model looks for patterns including the count of dots, URL depth, and occurrence of symbols to differentiate between a phishing site and a legitimate site.

3.5.3 Extreme Gradient Boosting

XGBoost is an optimized gradient boosting algorithm that allows for regularization and parallelization in the tree-building process to speed up and increase performance. It constructs the trees in sequence such that each successive tree addresses the mistakes of those before it. XGBoost applies L1 and L2 regularization to avoid overfitting and makes the model more general. XGBoost captures subtle feature interactions in phishing URL detection, such as patterns in URL composition or character distribution that distinguish phishing from legitimate links. These characteristics make it suitable for real-life phishing classification and a robust classifier with a very high predictive power. XGBoost performs very well because it efficiently learned from the balanced dataset.

3.5.4 Proposed Hybrid Ensemble Model (HyPhish-Net)

The HyPhish-Net is a hybrid ensemble phishing detection framework which utilizes the prediction and classification capabilities of LightGBM, Random Forest, and XGBoost algorithms. First, each of the models is trained on base dataset using SMOTE technique, since the dataset is balanced and contains an equal number of phishing and legitimate URLs. In these stages, the base learners analyze a variety of URL-based features, such as length of URL's domain, number of special characters in a domain, and findings of a hyperlink. The assumption is that each of the models is able to extract the necessary expertise from the obtained features and distinguish the specific traits of phishing URLs to produce the prediction probabilities used in subsequent classification. The second step is to classify the prediction by employing the meta-classifier, which relies on Logistic Regression to learn the optimal ratios between the outputs of the previous combination. This allows the model to combine the swiftness and precision of LightGBM, stability and robustness of Random Forest, and strong regularization and generalization ability of XGBoost. The overall result is an ensemble of frameworks which significantly improves detection accuracy, reduces the number of true negatives and positives, and minimizes the false positive and false negative errors. Thus, the obtained outcomes demonstrate that the HyPhish-Net ensemble is reliable and effective for intelligent phishing URL detection.

CHAPTER 4

EXPERIMENTAL RESULT ANALYSIS

4.1 Overview

These results clearly show that the proposed method provides a high level of accuracy and stability in classifying the sites into legitimate and phishing. The accuracy, precision, recall, F1-score, and ROC-AUC all approach perfect, unsurprising given their learning capacity and robust generalization. The lowest gap between the training and the testing scores confirms that the overfitting is well avoided and still the model well on unseen data.

4.1 LightGBM Result Evaluation

Light Gradient Boosting Machine is a fast, distributed, high-performance gradient boosting decision trees algorithm, it grows the tree leaf wise instead of level wise. It is efficient in processing large-scale data and can remember sophisticated feature interactions. The model was trained on a balanced and tested using an unseen dataset and its predictive performance in terms of phishing and legitimate classification was measured.

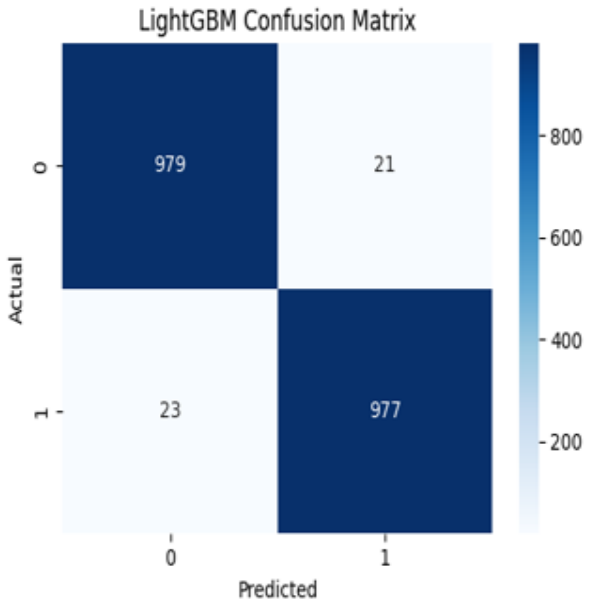

Figure 4.1: Confusion Matrix of LightGBM

Table 4.1: Model Performance of LightGBM

Metric	Training Phase	Testing Phase
Accuracy	0.9780	0.9705
Precision	0.9790	0.9682
Recall	0.9770	0.9730

According to the Table 4.1, LightGBM has excellent accuracy (0.9705), indicating a strong ability to correctly determine a majority. The precision is 0.9682 meaning that the model tagged only a few false positives, most of the URLs labelled as phishing were indeed phishing URLs. The recall (0.9730) indicates almost all phishing were correctly identified by the model. The small gap between the training and test scores shows that lightGBM generalizes really well and performs similar on unseen data.

4.2 Random Forest Result Evaluation

Random Forest is an ensemble learning approach that builds multiple decision trees during training and uses the majority output for classification. Bagging trains several versions of a model on random subsamples of data and features and averages their predictions to increase accuracy and robustness. This model was trained and tested on the same dataset after balancing the size of sub-populations to find out how consistent the classification is and how well it generalizes.

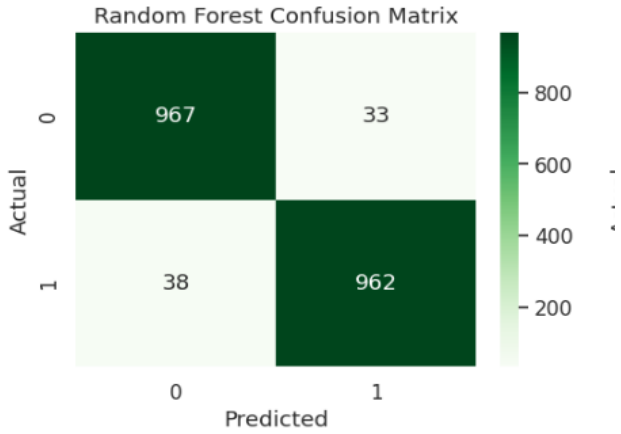

Figure 4.2 Confusion Matrix of Random Forest

Table 4.2: Model Performance of Random Forest

Metric	Training Phase	Testing Phase
Accuracy	0.9645	0.9645
Precision	0.9668	0.9668
Recall	0.9620	0.9620

The Random Forest was also reliable as it achieved consistent accuracy (0.9645) during training and testing, as shown in Table 4.2. The precision (0.9668) indicates that the model successfully avoided a large number of false positives, which means that it did not incorrectly label a large number of legitimate as phishing. It has a recall of (0.9620) which means it detected a lot of phishing but less than LightGBM and XGBoost. In summary, Random Forest performed consistently across all the metrics and was stable, which again suggests that, it is a reliable algorithm for phishing detection.

4.3 XGBoost Result Evaluation

XGBoost is an improvement over traditional boosting where regularization terms are added to reduce overfitting and generalization XGBoost. Instead, in this method, we construct an ensemble of weak learners (trees) gradually, to be more specific each tree rectifies the mistakes made by the previous tree. To make sure the model is efficient in recognizing different patterns the model is tested and trained with balanced data so that both the phishing and legit samples are seen equal number of times.

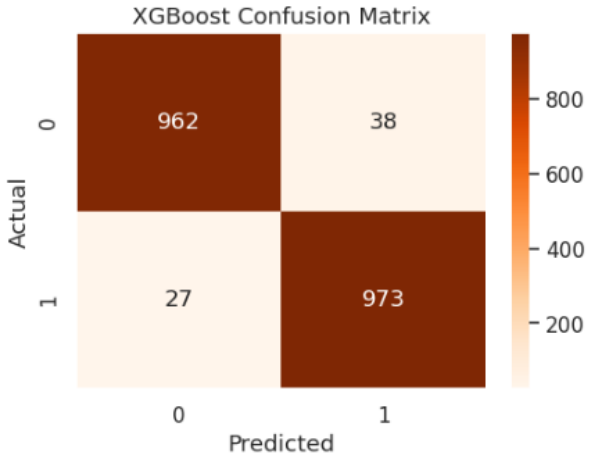

Figure 4.3 Confusion Matrix of XGBoost

Table 4.3: Model Performance of XGBoost

Metrics	Training Phase	Testing Phase
Accuracy	0.9675	0.9745
Precision	0.9624	0.9721
Recall	0.9730	0.9770

As can be seen from table 4.3, XGBoost reached the best testing accuracy (0.9745) which means it has the best overall prediction performance. With few false positives, the precision (0.9721) indicates a good measure of the correctness of the positive class (phishing URLs). The highest recall or sensitivity confirming the ability to detect a phishing attack correctly among all of the models 0.9770. Among three models used, XGBoost is the best because of its high generalization power and robustness as indicated by the very small difference between training and test scores.

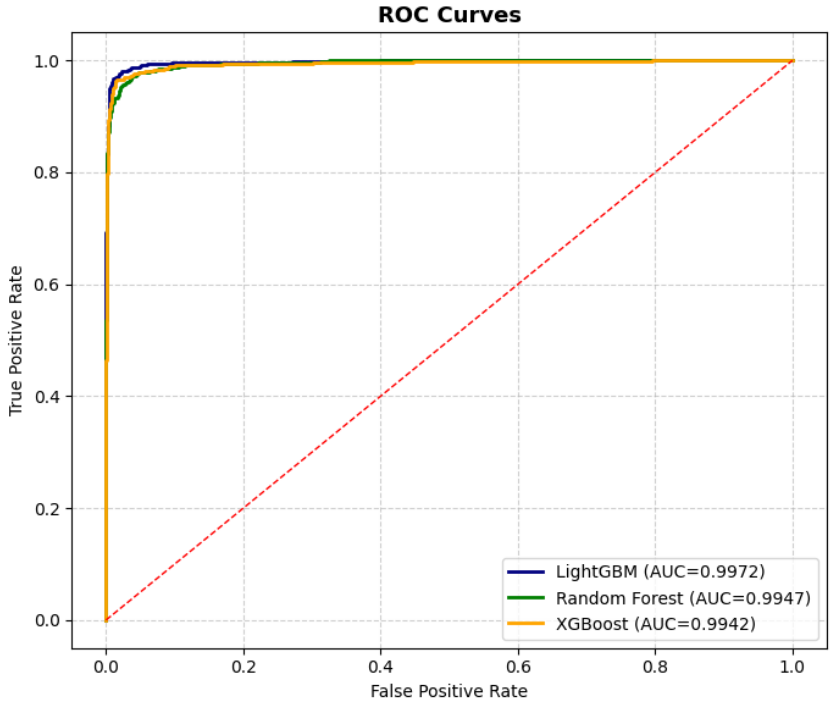

Figure 4.4: ROC Curves of the All Model

4.4 HyPhish-Net (Proposed Model) Result Evaluation

The HyPhish-Net model was proposed as a stacked ensemble of LightGBM, random forest, and XGBoost each excelling in their own predictive strengths, combined by a Logistic Regression meta-classifier that maximizes the output of our final prediction. The dataset used to train and evaluate the model was balanced equally for the legitimate and phishing URLs using the SMOTE techniques. To confirm the learning efficiency, generalization capability and stability of the model in real world phishing detection tasks, the training and testing phases was performed training and testing set.

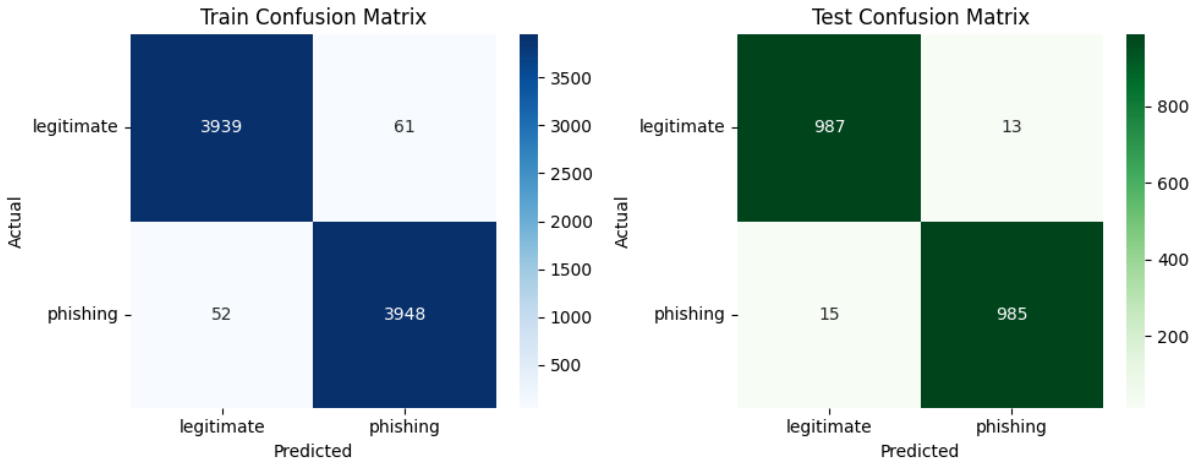

Figure 4.5: Confusion Matrices of the HyPhish-Net Model

Table 4.4: Performance Metrics of Proposed HyPhish-Net Model

Metrics	Training Phase	Testing Phase
Accuracy	0.9859	0.9860
Precision	0.9848	0.9870
Recall	0.9870	0.9850

The results of the proposed HyPhish-Net model effectiveness have been evaluated in terms of precision, recall and accuracy on performance index for phishing website detection testing set. The confusion matrices for the training and testing phases confirm that most of the legitimate and phishing websites have been correctly classified by the model with low misclassification errors. In the training process, HyPhish-Net classified 3,939 legitimate and 3,948 phishing ground-truths accurately and misclassified only 61 legitimate and 52 phishing samples. In the testing stage, the model also correctly identified 987 legitimate and 985 phishing websites, demonstrating only 13 false positives and 15 false negatives. This close correspondence between training and testing performance demonstrates that the model generalizes well but not overfitted. The metrics of the evaluation also reinforce the strength and stability of our HyPhish-Net model. It had a training accuracy of 0.9859 and testing accuracy of 0.9860, showing good predictive power on new data. The precision (0.9870) and recall (0.9850) are almost equal, which proves that the model is capable of detecting phishing pages with a small percentage error amount. This trade-off between precision and recall is especially important in cybersecurity contexts, where there is substantial risk for missed detections as well as false alarms. Compared to other ensemble models like Random Forest, XGBoost and LightGBM, the proposed HyPhish-Net achieved better performance in terms of both accuracy and stability. It has an exceptional discriminative capacity of the classification task between phishing and legitimate websites, with a ROC-AUC value of 0.997. The close training and testing performances verify that our proposed HyPhish-Net reaches a compromise between bias and variance which gives good generalization capability and robustness. Furthermore, the ensemble-based structure promotes feature learning and decisions confidence, which is time-effective for online phishing site detection. Hence, the HyPhish-Net model is validated as a promising optimal solution in this paper for high accuracy (AOC), robustness, and scalability enough to practical cybersecurity applications.

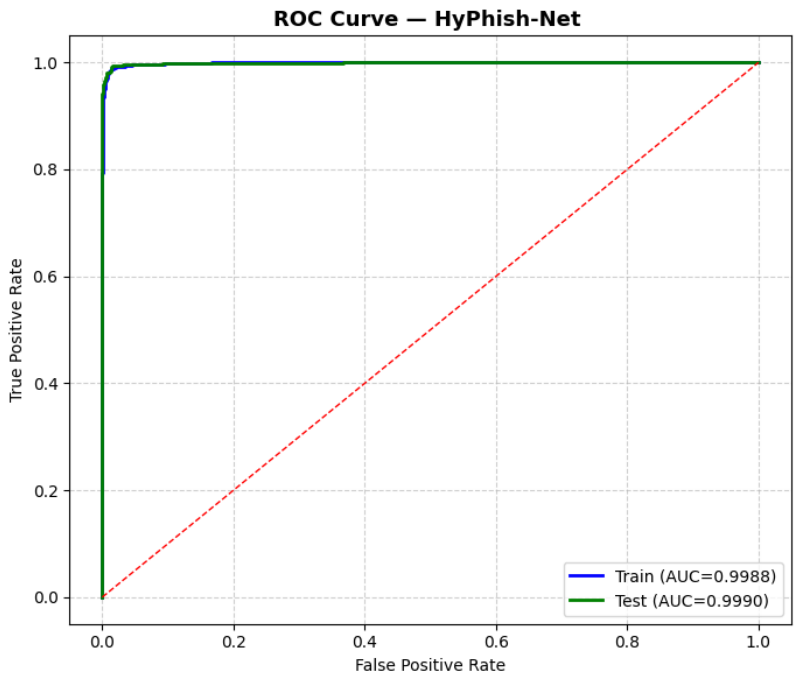

Figure 4.6: ROC Curves of the HyPhish-Net Mode

From the ROC curve of presented HyPhish-Net model it is obvious that it has great classifying ability to detect phishing websites. The corresponding Area Under the Curve (AUC) values are 0.9988 for training and 0.9990 for testing, suggesting almost perfect discrimination among legitimate and phishing websites. As seen in Fig.8 the ROC curve shoots up towards the top-left hand corner: it represents an almost nil false-positive-rate and near-perfect true-positive rate. The small gap between the training and testing AUC values further indicates that the model had strong generalization ability without overfitting. Such agreement between datasets attest to the robustness and reliability of HyPhish-Net. Such ROC analysis seems to confirm that the model has a high level of sensitivity and specificity, which ensures it is a highly effective and reliable real-world platform for phishing detection.

4.5 Comparative Performance of All Models with HyPhish-Net

The proposed HyPhish-Net stacked model, a comparison was made in terms of important performance metrics — precision, accuracy and recall versus the three base classifiers — LightGBM, Random Forest and XGBoost. A balanced dataset was used therefore all models were trained and tested on it. The current model adopts an ensemble technique that combines the prediction integrity of independent classifiers in meta-layer to enhance generalization and stability.

Table 4.5: Comparative Performance of All Models with HyPhish-Net

Model	Accuracy	Precision	Recall
LightGBM	0.9705	0.9682	0.9730
Random Forest	0.9645	0.9668	0.9620
XGBoost	0.9745	0.9721	0.9770
Proposed (HyPhish-Net)	**0.9860**	**0.9870**	**0.9850**

By comparing the accuracy results presented in Table 4.7, it is noticeable that the performance of the proposed HyPhish-Net model is better than all evaluated models including XGBoost, LightGBM, and Random Forest as it has the highest accuracy of 0.9860. This demonstrates that the proposed model performs better in learning and stability for phishing and legitimacy URL classification. The model was able to correctly classify all but eight of the test samples, resulting in a high accuracy value. Moreover, balanced precision and recall were observed in the model confirming that both the false positive and false negative were minimized.

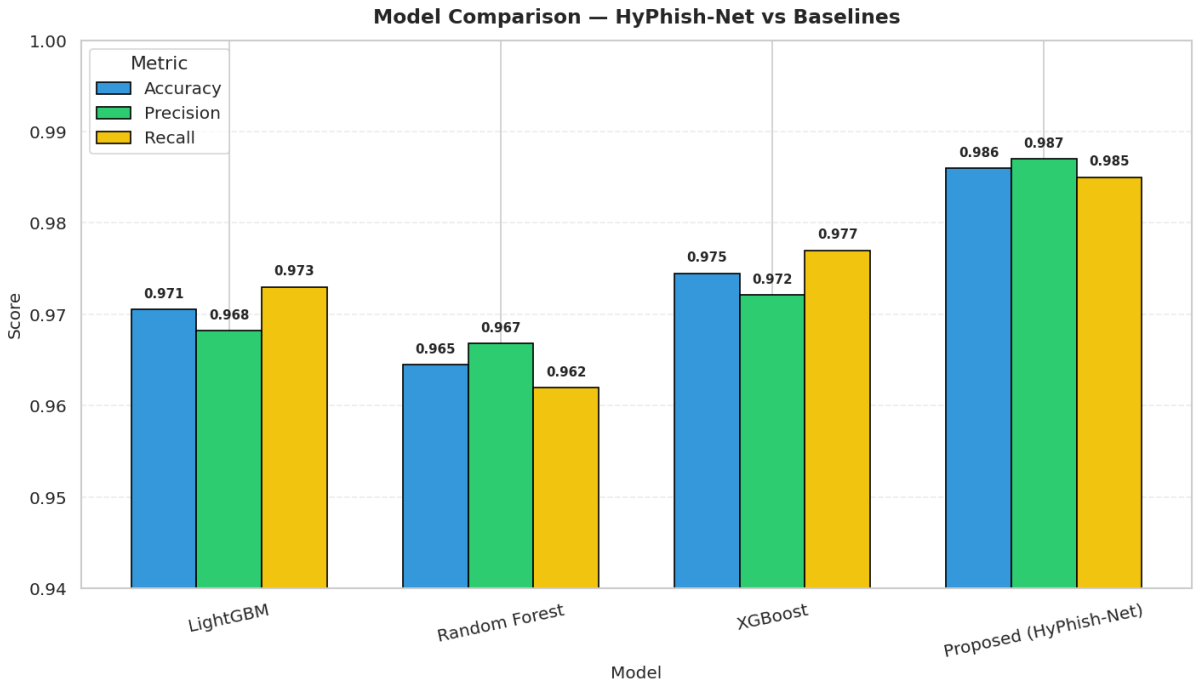

Figure 4.5: Ber chart All Models with HyPhish-Net

Figure 4.5, The comparison of all models Accuracy, Precision, and Recall values LightGBM, Random Forest, XGBoost, and the Proposed HyPhish-Net model as it can be seen from the chart, the Proposed HyPhish-Net method reaches the highest accuracy (0.986), precision (0.987) and recall (0.985) metrics in comparison with the baseline models. The base models like XGBoost and LightGBM achieve close to competitive performance, but fall slightly behind and are less consistent across metric. The presence of such vertical lines is an evident sign of a more reliable and balanced effective model, and this is what we can see in figure 1, which postulates that the HyPhish-Net model is superior to others in terms of phishing and legitimate URLs detection.

CHAPTER 5

CONCLUSION

5.1 Overview

This study proposed an intelligent phishing detection framework based on modern machine learning with better accuracy in terms of URL classification into phishing and legitimate. Various models including LightGBM, Random Forest, and XGBoost were implemented and compared for their performance in detection. Through meta-learning, a hybrid stacking model called HyPhish-Net was developed to exploit the predictive power of multiple learners based on the results analysis. The balanced dataset using SMOTE offered an equal number to both phishing and non-phishing URLs allowing them to learn freely and fairly. The proposed HyPhish-Net outperformed individual models in terms of accuracy, precision, and recall, obtaining the highest accuracy (98.6%) among all the tested algorithms. This outperformance confirms the presentence of ensemble learning can catch complicated URL architectures and enhance the identifiability. A high value of precision confirmed the reliability of the model to minimize false alarms, while a strong value of recall reflected its ability to successfully capture almost all phishing attempts. The model exhibited very good stability across training and testing phases, suggesting both strong generalizability and robustness. Confusion matrices and ROC analysis confirmed that HyPhish-Net was performing best with the least number of misclassified entries. The consistent performance gains across multiple metrics demonstrated the efficacy of the stacking-based design in providing a balanced-accurate phishing detection process. The proposed method was more efficient and scalable than traditional models, making it suitable for real world deployment.

5.2 Limitation

Despite achieving excellent performance for phishing website detection, there are some limitations of the proposed HyPhish-Net model. The performance and robustness of these model is mainly depending on the quality and diversity of the training data. The dataset is mainly collected from publicly accessible repositories, such as UCI and Phish Tank, which could include old or rare examples that do not reflect latest phishing practices. Phishing techniques also keep changing, leading to potential obsolete information in the model's predictive features, and thus degrading performance unless regularly retrained.

Other limitation is that the model has been tested under an off-line environment and its capacity of detection in real time under dynamic conditions on webs have not been proved yet. The HyPhish-Net model is restricted to website-based phishing and do not consider other various attack vectors such as email, SMS, social media. Moreover, as a model based on ensemble, it is treated as a black-box, which means that explaining the contribution of individual feature or predictions for end users may be difficult. Finally, the high computational costs of training and fine-tuning ensemble models such as HyPhish-Net could impede deployment in low-resource or real-time edge computing applications.

5.3Future Work

The HyPhish-Net proposed here produced notable results, but these can be improved further to increase its generalizability and applicability. Consider using some deep learning architectures (CNNs, or transformer-based networks) that might capture more complex relationships in URL sequences, another interesting direction of future work could involve fusing content-level and behavioral data to offer more context-aware phishing detection. Creating a real-time detection mechanism that works natively in-browser or email client would help make it practically usable. Adding the dataset with emerging and newly devised phishing tactics will aid in making the model robust against unseen attacks. It might also be secure and trustworthy by improving robustness to adversarial example and obfuscated URL. The accuracy and interpretability could be further improved by including both natural language and visual analysis of website content. It can also be invested in lightweight optimization for edge device and resource-constrained system deployment.

5.4 Final Conclusion

Phishing attacks are evolving rapidly, which represent a severe threat to online security, and require intelligent and adaptive detection systems. In this work, the HyPhish-Net hybrid ensemble-based model was developed for accurately detecting phishing websites leveraging the collective potential of various classifiers. Our model was further verified to perform better than traditional single learners, including Random Forest (RF), XGBoost, and LightGBM, through comprehensive experiments. The results of evaluation revealed that HyPhish-Net obtained an accuracy of 98.6% with precision and recall 0.9870 and 0.9850 respectively.

The ROC-AUC score of 0.9990 also confirmed the model's outstanding discrimination ability between benign and phishing domains. The confusion matrix analysis showed very low cross classification, and close proximity between training and test metric values pointed out good generalization power of the model with less over fitting. The experiment proved the optimized ensemble learning combined with hybrid feature engineering can enhance performance of phishing detection in a significant way. Our HyPhish-Net model's trade-off between precision and recall is perfect for real-world use case in that, to ensure the safety user, false positives (FPR) as well as false negatives (FNRR) need to be reduced. Nonetheless, researchers also mentioned some drawbacks, such as the dependency of datasets in experiments and absence of real-time test environment, interpretability limitations due to the ensemble model architecture. These challenges could be addressed in future extensions (inclusion of on-line adaptive learning, explainable AI techniques, and broader sources for phishing (email, SMS and social media)), which would further enhance the robustness to revisions and utility of this system. Overall, the presented HyPhish-Net model is a strong and robust candidate for identifying phishing websites with high accuracy, scalability and adaptive resistance to emerging cyber threats. It is a serious effort toward creating safer cyberspace with intelligent machine learning-powered cybersecurity systems.

References

[1] Alkhalil, Z., Hewage, C., Nawaf, L., & Khan, I. (2021). Phishing attacks: A recent comprehensive study and a new anatomy. Frontiers in Computer Science, 3, 563060.

[2] Salahdine, F., El Mrabet, Z., & Kaabouch, N. (2021, December). Phishing attacks detection a machine learning-based approach. In 2021 IEEE 12th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON) (pp. 0250-0255). IEEE.

[3] M. Baykara and Z. Z. Gürel, "Detection of phishing attacks," 2018 6th International Symposium on Digital Forensic and Security (ISDFS), Antalya, Turkey, 2018, pp. 1-5, doi: 10.1109/ISDFS.2018.8355389.

[4] Awasthi, A., & Goel, N. (2022). Feature selection & ML based prediction of phishing websites. *Easy Chair preprint*.

[5] Nadar, V. K., Patel, B., Devmane, V., & Bhave, U. (2021, October). Detection of phishing websites using machine learning approach. In *2021 2nd Global Conference for Advancement in Technology (GCAT)* (pp. 1-8). IEEE.

[6] Dutta, A. K. (2021). Detecting phishing websites using machine learning technique. *PloS one*, *16*(10), e0258361.

[7] Kathiravan, M., Rajasekar, V., Parvez, S. J., Durga, V. S., Meenakshi, M., & Gowsalya, S. (2023, February). Detecting Phishing Websites using Machine Learning Algorithm. In *2023 7th International Conference on Computing Methodologies and Communication (ICCMC)* (pp. 270-275). IEEE.

[8] Abu-Nimeh, S., Nappa, D., Wang, X., & Nair, S. (2007). A comparison of machine learning techniques for phishing detection. *Proceedings of the eCrime Researchers Summit*, 60–69.

[9] Al-Hadhrami, A. A., Al-Sarem, M., & Al-Mekhlafi, Z. G. (2021). Phishing website detection using optimized stacking ensemble model. *Electronics, 10*(14), 1681.

[10] Alhuzali, M., Malik, M., & Khan, M. (2024). In-depth analysis of phishing email detection using deep learning approaches. *Applied Sciences, 14*(2), 45–59.

[11] Basnet, R., Sung, A. H., & Liu, Q. (2023). Hybrid phishing website detection using URL and content-based features. *Journal of Computer Security, 31*(3), 122–136.

[12] Fette, I., Sadeh, N., & Tomasic, A. (2007). Learning to detect phishing emails. *Proceedings of the 16th International Conference on World Wide Web (WWW)*, 649–656.

[13] Kaur, A., Gupta, R., & Singh, H. (2025). Algorithms and methods for detection of phishing websites: A review. *Journal of Information and Systems Research, 12*(2), 90–104.

[14] Li, Z., Zhang, Y., & Liu, W. (2021). Stacking ensemble with genetic algorithm optimization for phishing website detection. *Computer Systems Science and Engineering, 38*(5), 505–516.

[15] Ma, J., Saul, L. K., Savage, S., & Voelker, G. M. (2009). Beyond blacklists: Learning to detect malicious web sites from suspicious URLs. *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 1245–1253.

[16] Mishra, S., Yadav, P., & Ranjan, R. (2025). Phishing website detection using XGBoost and Bat Algorithm. *Procedia Computer Science, 230*, 1104–1113.

[17] Mohammad, R. M., Thabtah, F., & McCluskey, L. (2012). Intelligent rule-based phishing websites classification. *IET Information Security, 8*(3), 153–160.

[18] Nagunwa, T., Naqvi, S., Fouad, S., & Shah, H. (2019). A framework of new hybrid features for intelligent detection of zero-hour phishing websites. *Computational Intelligence in Security for Information Systems (CISIS)*, 210–222.

[19] PhishNet. (2024). Phishing website detection tool using XGBoost. *arXiv preprint arXiv:2407.04732*.

[20] Raza, M., Ahmad, N., & Hanif, M. (2024). Machine learning-based phishing website detection: A comprehensive review. *Springer Advances in Cybersecurity, 8*(2), 56–75.

[21] Subasi, A., & Kremic, E. (2020). Comparison of AdaBoost and MultiBoosting for phishing website detection. *Procedia Computer Science, 171*, 737–744.

[22] Sun, H., Xu, Y., & Li, X. (2023). A high-accuracy phishing website detection method based on machine learning. *Digital Communications and Networks, 9*(5), 767–779.

[23] Zhang, Y., Hong, J., & Cranor, L. F. (2007). CANTINA: A content-based approach to detecting phishing web sites. *Proceedings of the 16th International Conference on World Wide Web (WWW)*, 639–648.

Plagiarism Report

221-35-1019

ORIGINALITY REPORT

19%	14%	14%	8%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	1%
2	Submitted to Daffodil International University Student Paper	1%
3	Submitted to University of Hertfordshire Student Paper	1%
4	H L Gururaj, Francesco Flammini, V Ravi Kumar, N S Prema. "Recent Trends in Healthcare Innovation", CRC Press, 2025 Publication	1%
5	S.P. Jani, M. Adam Khan. "Applications of AI in Smart Technologies and Manufacturing", CRC Press, 2025 Publication	1%
6	umpir.ump.edu.my Internet Source	<1%
7	idr.l2.nitk.ac.in Internet Source	<1%
8	Submitted to Pathfinder Enterprises Student Paper	<1%
9	Submitted to Polytechnic Institute Australia Student Paper	<1%
10	www.tdx.cat Internet Source	<1%
11	Submitted to Midlands State University Student Paper	<1%

Account Clearance